

HCV Quasispecies Evolution: Association with Progression to End-Stage Liver Disease in Hemophiliacs Infected with HCV or HCV/HIV

**Hongxing Qin¹, Norah J. Shire¹, Erica D. Keenan¹, Susan D. Rouster¹,
M. Elaine Eyster², James J. Goedert³, Margaret James Koziel⁴,
Kenneth E. Sherman¹, and the Multicenter Hemophilia Cohort Study Group**

¹University of Cincinnati College of Medicine, Cincinnati, OH; ²Pennsylvania State University, Hershey, PA; ³Viral Epidemiology Branch, National Cancer Institute, Rockville, MD; ⁴Beth Israel-Deaconess Medical Center, Harvard University, Boston, MA

Keywords: Hemophilia, HCV, HIV, quasispecies, evolution

Running Title: HCV EVOLUTION IN HEMOPHILIACS

Correspondence: Kenneth E Sherman, MD, PhD
Division of Digestive Diseases
Univ. of Cincinnati College of Medicine
231 Albert Sabin Way
Cincinnati, OH 45267-0595
Phone: 513-558-7200
Fax: 513-558-1744
e-mail Kenneth.Sherman@uc.edu

Funding for this study was provided by the National Institute of Allergy and Infectious Diseases, NIH (grant number R01 AI49508) and the U.S.P.H.S. General Clinical Research Centers Program, National Center for Research Resources, NIH (grant numbers M01 RR010732, M01 RR08084, C06 RR016499).

Word Count: Abstract 189
Text 3845

Scientific Heading: Clinical Observations, Interventions, and Therapeutic Trials

Abstract

Patients with inherited bleeding disorders who received clotting factor concentrates before 1987 have high rates of HCV or HCV/HIV infection. We evaluated HCV quasispecies evolution in longitudinally collected specimens comparing those from patients with progression to end-stage liver disease (ESLD) (cases) to those with compensated chronic hepatitis (controls). Plasma samples were obtained from the NCI Multicenter Hemophilia Cohort Study. Controls were matched for age, gender, infection duration, and presence/absence of HIV. Samples from early infection were compared to those obtained after onset of ESLD in the cases. The first hypervariable (HVR1) and core protein coding regions were amplified, subcloned and sequenced. Complexity and diversity were determined. Over 700 subclones from 10 pairs of patients (8 with HIV) followed over approximately 9.3 years were evaluated. HVR1 complexity narrowed over time in the cases, while it increased in controls ($p = 0.01$). Similar trends were observed for diversity within HVR1 and the core region ($p \leq 0.04$). HCV-infected patients with inherited bleeding disorders undergo quasispecies evolution over time. Evolution patterns differ for progressors and nonprogressors. Further understanding of these mechanisms may help identify factors related to progression rate and treatment response.

Introduction

Hepatitis C virus (HCV) infection is associated with progression to end-stage liver disease (ESLD) in a significant subset of chronically infected patients.

Hemophiliacs and other patients with inherited bleeding disorders who received clotting factor concentrates prior to 1987 are at particular risk for acquiring HCV infection from recurrent exposure to concentrates made from plasma pools prepared from 20,000 or more blood donors at a time when single-unit risk of infection approached 5%.¹ Through the 1980's the HCV infection risk was compounded by transmission of human immunodeficiency virus (HIV) in blood products. Increased rates of progression to ESLD in those with chronic HCV infection have been well documented in hemophiliacs and in other populations with HIV coinfection.²⁻⁴

HCV replicates at very high levels with production of virions estimated at up to 10^{12} copies/day.⁵ This high rate of replication, combined with lack of an error correction mechanism and ongoing but variable immune selection, results in development of both synonymous and nonsynonymous mutations. Immune selection puts pressure on key antigen-recognition sites and drives the emergence of a closely related virus family that can be isolated from the serum of infected patients. Previous evaluation of viral population dynamics has provided critical insights into short-term outcomes including early spontaneous viral clearance,⁶ interferon-associated viral clearance,^{7,8} and HCV emergence following liver transplantation.^{9,10} However, lack of suitable long-term longitudinal cohorts has limited study of the relationship between HCV quasispecies evolution and development of ESLD. Furthermore, most studies assessing quasispecies

heterogeneity have focused on the first hypervariable region (HVR1) within the E2/NS1 portion of the HCV genome. However, the core region has been associated with multiple functions, including regulation of tumor suppressor genes,¹¹⁻¹³ modification of cell susceptibility to apoptosis,^{14,15} and inhibition of HIV-1 and hepatitis B virus replication.^{16,17} Therefore, studies of the core region may provide important insights into the relationship between quasispecies and disease progression.

In this report, we evaluated the emergence of HCV quasispecies in a longitudinal cohort of hemophiliacs with either HCV or HCV/HIV coinfection among subjects with clear clinical progression to ESLD compared to emergence in matched controls without evidence of clinical progression.

Materials and Methods

Study Population

Plasma samples from patients with known HCV and HIV status were collected from the NCI Multicenter Hemophilia Cohort Study (MHCS) and stored at -80° C. This study cohort was initiated for the study of HIV infections in 1982 and included periodic clinical evaluation and testing for HIV antibodies beginning in 1984 and HCV antibodies beginning in 1990.⁴ All subjects provided informed consent at the time of enrollment, and the protocol was approved at Institutional Review Boards at each of the participating clinical sites. Laboratory evaluation included complete blood count, serum transaminases (aspartate aminotransferase (AST) and alanine aminotransferase (ALT)), hepatitis B antigen and antibody status, and CD4+ counts. These were collected approximately annually over a minimum of five years for two comparison groups of patients comprising cases and controls. Cases consisted of either HCV/HIV coinfecting or HCV monoinfected

patients with liver failure, defined as persistent ascites, variceal bleeding, hepatic encephalopathy or death attributed to liver disease.^{3,4} The control group included coinfecting or monoinfecting patients without liver failure and with stable liver function tests. Cases and controls were matched by age, gender, HIV status, and estimated duration of HCV infection at first and last samples tested herein. Dates of initial HCV infection were estimated using history of use of clotting factor concentrate (first use = infection) and plasma or cryoprecipitate (midpoint between first plasma/cryoprecipitate and first concentrate use or 1990). There were 10 matched pairs overall: 8 pairs were HCV/HIV coinfecting and 2 pairs were HCV monoinfecting. The preponderance of coinfecting pairs is representative of the proportional difference in coinfecting vs. monoinfecting patients with liver failure within the cohort.

HCV RNA isolation, HVR1 and core region amplification

HCV RNA was extracted from 140 µl of patient plasma using a QIAamp Viral RNA Mini kit (QIAGEN, Valencia, CA) according to the manufacturer's protocol. Amplification of the E2/NS1 was carried out using a nested RT-PCR method described by Sherman et al.¹⁸ The first set of primers used were: external forward (EF), position 1048 – 1067, 5'-GGTGCTCACTGGGGAGTCCT-3'; external reverse (ER), position 1291c – 1269c, 5'-CATTGCAGTTCAGGGCCGTGCTA-3'; internal forward (IF), position 1087 – 1106, 5'-TCCATGGTGGGGAACTGGGC-3'; and internal reverse (IR), position 1262c – 1243c, 5'-TGCCAACTGCCGTTGGTGTT-3', with the primer positions corresponding to the HCV sequence described by Choo et al.^{19,20} These primers flank a 136-bp region of the E1/E2 envelope coding domain that includes the HVR1.

The second set of primers, used to amplify a 1270-bp product which includes the HCV core protein, were: EF, position –62 to –42, 5'-CTTGTGGTACTGCCTGATAGG-3'; ER position 1291c – 1269c, 5'-CATTGCAGTTCAGGGCCGTGCTA-3'; IF, position –51 to –29, 5'-GCCTGATAGGGTGCTTGCGAGTG-3', and IR, position 1262c – 1243c, 5'-TGCCAACTGCCGTTGGTGTT-3'.

Cloning and Sequencing of cDNA

The HCV cDNA products were ligated into pGEM-T Easy Vector and used to transform JM109 high efficiency competent cells according to the manufacturer's protocol (Promega Corp., Madison, WI). Each transformation culture was plated out on LB/ampicillin/IPTG/X-Gal plates and grown overnight at 37 °C. White colonies containing the inserts were selected and grown overnight in 5 ml of LB broth containing ampicillin in a shaking incubator at 37 °C. The plasmid DNA was then purified using the QIAprep Spin Miniprep kit (Qiagen, Valencia, CA) and the presence of the inserts confirmed by digestion with EcoR1. Sequencing was performed on approximately ten clones per patient at the early (first available) and late (first available after progression to ESLD) timepoints and their matched control samples. Due to limited sample availability, it was not possible to clone and sequence samples of matched pairs at interim timepoints.

DNA Sequence Manipulation and Phylogenetic Analysis

Sequences were aligned manually and gaps created where any additions or deletions occurred. Phylogenetic and molecular evolutionary analyses were conducted using Molecular Evolutionary Genetic Analysis software (MEGA version 2.1).²¹ The

MEGA algorithms were used to calculate the mean Kimura 2-parameter pairwise distance for all clones as well as a matrix of Kimura 2-parameter pairwise distances for each patient. The numbers of synonymous mutations (dS) and the numbers of nonsynonymous mutations (dN) were determined using the Nei-Gojobori method with the Jukes-Cantor correction for multiple substitutions.²² The mean dN/dS ratio, a measure of selective pressure, was also calculated for each patient. Phylogenetic trees of each patient, illustrating genetic distance between clonal sets, were generated using the Neighbor Joining method of Saitou and Nei.²³ Samples from early infection in both cases and controls were compared to those obtained at follow-up from cases and controls, after onset of ESLD in the cases.

Statistical Analysis

Demographic comparisons were performed using the chi-square test and the paired t-test. Nonlinear and linear regression models, the paired t-test, and a two-way mixed model ANOVA were used to assess differences in complexity and distance matrices between case/control pairs over time. These analyses were conducted for the 136-bp product amplified from the internal forward and internal reverse primers and the truncated 81-bp sequence corresponding to the HVR1 region, and for the core region from the core internal forward through 573 base pairs. The difference in amino acid diversity between pairs was evaluated using a paired t-test by amino acid position. Specific amino acid differences and hydrophobic character differences were evaluated using the chi-square test. Two-tailed p-values less than 0.05 were considered statistically

significant. Statistical analyses were conducted using either Stata version 7.0 (Stata Corporation, College Station, TX) or SAS version 8 (SAS Institute, Cary, NC).

Nucleotide Sequence Accession Numbers

Nucleotide sequences used in this study are available in the GenBank database under accession numbers AY366563-AY366891, AY367040-AY367049 for the hypervariable region and AY521893-AY522326 for the core region.

Results

Patient Demographics

Table 1 shows the demographic characteristics of the cases and controls. All patients were male, and all were HCV genotype 1. Nineteen out of 20 patients (95%) had prior, resolved coinfection with hepatitis B virus (HBV). One patient was HBV-marker-negative. Baseline CD4+ and CD8+ T cell counts, AST and ALT levels, and HCV/HIV RNA titers were not available for all patients. Baseline values were used when available; otherwise, the next measurement after baseline was utilized.

Analysis of the HVR1

HCV Quasispecies Complexity

A total of 342 clones from 20 patients were sequenced and analyzed. The mean time between early and late samples was 9.3 years (SE 0.69 years). Mean numbers of clones per patient were 9.9 (SE 0.76) for the early timepoint and 7.4 (SE 0.34) for the follow-up timepoint. Complexity was assessed by the proportion of unique nucleotide sequences to the number of clones assessed at each timepoint for each patient (Table 2). Regardless of HIV status, the mean early-timepoint frequency of unique nucleotide sequences within HVR1 for progressors was 0.73 (SE 0.04) sequence/clone for progressors, compared to 0.59 (SE 0.08) for non-progressors ($p = 0.23$). At the late timepoint, the mean frequencies were 0.61 (SE 0.04) and 0.77 (SE 0.07) for progressors vs. nonprogressors, respectively ($p = 0.22$). The differences between early and late frequencies indicate a narrowing of the quasispecies population complexity for progressors (-0.12 sequence/clone) and an increase in complexity (0.18 sequence/clone)

for nonprogressors. The changes in complexity over time were statistically significantly different for progressors compared to nonprogressors ($p = 0.01$).

HCV Quasispecies Diversity

Diversity was assessed by comparing the mean Kimura 2-parameter distances between early and late sets of clones for matched cases/controls. Early and late distances are displayed in Figure 1. Regardless of HIV status, progressors as a group tended to have greater genetic diversity at the early timepoint than nonprogressors ($p = 0.04$) and demonstrated significant narrowing of genetic diversity over time ($p = 0.04$). Nonprogressors as a group maintained the same level of diversity over the study duration. MEGA was used to compute net change in intrasubject distances between early and late groups of clones by the formula $d_A = d_{XY} - (d_X + d_Y)/2$ where d_{XY} is the average distance between groups X (early clones) and Y (late clones), and d_X and d_Y are the mean within-group distances. The mean intrasubject change in distance for progressors was 0.32 (SE 0.11) vs. 0.11 (SE 0.03) for controls ($p = 0.09$). Phylogenetic trees illustrating the distances between clones for each patient are shown in Figure 2.

Mutational Selection

The nonsynonymous (dN) to synonymous (dS) mutation ratio, a marker of selective pressure, was compared between early and late time points for both cases and controls. One patient (pair 2 case), who was infected with two HCV subtypes, displayed a dramatically large dN/dS ratio at the later timepoint (39.02) due to a shift in the dominant subtype from 1a to 1b. With this outlying matched pair removed from the analysis,

progressors demonstrated an overall statistically non-significant decrease in dN/dS ratio (from 2.02 to 0.62, $p = 0.13$) over time while nonprogressors remained stable (Table 3).

Amino acid sequences were predicted for the 81-bp HVR1 (aa 384-410) and aligned to determine the number of unique amino acids observed at each position. Sequences for each matched pair are shown in Figure 3. From these data, a profile of amino acid variation in HVR1 for cases and controls was created (Figure 4). The mean number of amino acid substitutions per site, including deletions, was 5.3 (SE 0.54) for the progressors and 5 (SE 0.61) for nonprogressors ($p = 0.44$). However, site-by-site analysis revealed statistically significant differences in the frequency of the most-conserved amino acid residues at various positions. A Thr residue is found at position 2 of the amplified sequence in all of the controls (151/151, 100%), but one progressor had a deletion at that location in 15/16 early-timepoint clones so that the frequency in the case group was 136/150 (91%). At position 4, Val is predominant in both cases and controls, but in 94/150 cases (63%) vs. 116/151 controls (77%, $p = 0.03$). Instead of Val, Thr occupies this position in 49/150 cases (33%) vs. 6/151 controls (4%). An Ala residue is present at position 9 in 84/151 control clones (56%), but only in 59/150 cases (39%, $p = 0.02$). Sixty-seven of 150 cases (45%) have a Val residue occupying that position vs. 13/151 controls (9%). Similar differences in frequencies of Thr and Asn are found at positions 13 and 27, respectively. Other positions are very highly conserved in both groups, including Gly at positions 6, 7, and 23, and Gln at position 2.

Because amino acid variation may alter hydropathic properties that are required for HVR1 structure and function, we assessed the proportions of hydrophilic, neutral, and hydrophobic residues at each site within the HVR1 using the scale of Black and Mould.²⁴ Overall, sites 2, 6, 7, 18, and 23 were at least 80% neutral in sequences from both cases and controls. Sites 25, 26, and 27 were nearly 100% hydrophilic. Sites 16, 19, 20, and 24 were nearly 100% hydrophobic. The remaining sites (1, 3, 4, 8-15, 17, 21, and 22) appeared more tolerant of variability. Ten sites had greater than 14 % difference ($p < 0.05$) in frequency for the predominant hydropathic character of its residues between sequences from cases and controls. This subset of ten sites is displayed in Figure 5.

Analysis of the Core Region

HCV Quasispecies Complexity

A total of 425 clones from the core protein coding region of the 20 patient pairs were sequenced and analyzed. The mean time between early and late samples was 9.3 years (SE 0.69 years). Mean numbers of clones per patient were 10.9 (SE 0.42) for the early timepoint and 10.35 (SE 0.29) for the follow-up timepoint. Complexity was assessed by the proportion of the number of unique core region nucleotide sequences to the number of clones from patients at early and late timepoints (Table 4). The total early-timepoint mean of unique nucleotide sequences per total number of clones within the core region for progressors, regardless of infection status, was 0.54 (SE 0.09). For nonprogressors, this mean was 0.76 (SE 0.05) ($p = 0.06$). At the late timepoint, the mean frequencies were 0.67 (SE 0.09) and 0.75 (SE 0.06) for progressors vs. nonprogressors, respectively ($p = 0.47$). In contrast to the changes in quasispecies population complexity

in the HVR1, where a narrowing of the population was seen in progressors compared to increased complexity for nonprogressors, no significant differences were apparent in the unique sequences of the core region for progressors vs. nonprogressors.

HCV Quasispecies Diversity

Diversity in the core region was assessed via Kimura 2-parameter pairwise distances between early and late sets of clones for the matched cases and controls. Differences in genetic distance are shown in Figure 6. The core region demonstrated less overall genetic diversity than the HVR1, as indicated by smaller Kimura 2-parameter distances. At baseline, there were no statistically significant differences in diversity between progressors and nonprogressors. Over time, 5 progressors and 1 nonprogressor showed a decrease in genetic distance within early and late sets of clones. Five progressors and 6 nonprogressors displayed an increase in diversity, and 0 progressors vs. 3 nonprogressors demonstrated no net change. However, mean between-group genetic distances for early and late clonal sets were 0.04 (SE 0.01) and 0.02 (SE 0.003) for progressors vs. nonprogressors, respectively ($p = 0.02$, data not shown). 80% of the progressors had greater overall genetic distance than their matched controls.

Mutational Selection

Mutational selection for the core region was determined by calculating mean dN/dS ratios for cases and controls over the early and late timepoints (Table 5). Although the progressor in pair 2 was clearly an outlier in the HVR1 dN/dS ratio analysis due to the significant nucleotide changes associated with a subtype shift, there were no outliers

evident in analysis of the core region. There was a statistically significant baseline difference in paired dN/dS ratio between progressors and nonprogressors. However, all dN/dS ratios were ≤ 1 , indicating a lack of selective pressure driving nonsynonymous mutations. Mean dN/dS ratios of both progressors and nonprogressors decreased over time so that late-timepoint differences were no longer significant.

Finally, no measures of quasispecies complexity, diversity, or mutational selection were found to be significantly associated with demographic characteristics, baseline CD4+/CD8+ counts, HCV viral load, or HIV infection status.

Discussion

Although it has been well documented that patients coinfectd with HCV/HIV are at increased risk for progression to end-stage liver disease compared to monoinfected patients,^{2-4,25-27} the evolution of the HCV quasispecies during this process has not yet been clearly defined. HCV quasispecies complexity and diversity have been examined in relation to HIV-related immunosuppression,²⁸⁻³⁰ patient response to interferon-based treatment,^{8,31-36} response to liver transplantation,^{9,37-40} and in patients with alcoholism^{18,41} or hemophilia.⁴² These studies have generated conflicting results. While some investigators have observed an association between complexity of the HVR1 and response to interferon,^{43,44} others have not.⁴⁵⁻⁴⁷ There is a paucity of longitudinal data regarding HCV quasispecies emergence and selection in patients progressing to ESLD. One report from Curran et al.⁴⁸ analyzed serum and biopsy specimens from six untreated HCV-infected patients with various stages of liver disease over approximately 5 years. Three patients with severe liver disease demonstrated a stable population while three patients with mild liver disease demonstrated increasing viral complexity. However, not all groups have detected such a difference.⁴⁹ This may be related to differences in sample size and duration of observation.

Studies in patients with inherited bleeding disorders have been similarly limited in scope and have yielded conflicting results. Many have focused on methods to detect the multiple genotypes, subtypes, and variants that may have existed in the donor pool rather than on the relationship between quasispecies heterogeneity and disease progression.^{50,51} Our study is the first to analyze HVR1 and core region variability in untreated HCV-

infected and HCV/HIV-coinfected hemophilic patients with progression to ESLD, compared to matched, nonprogressing controls.

In this study, we demonstrated significant differences in changes in quasispecies complexity over time within the HVR1 for progressors compared to nonprogressors ($p = 0.01$). Progressors had greater quasispecies complexity at baseline (more unique nucleotide sequences per number of clones), but this complexity narrowed over the follow-up period of 9 years. Nonprogressors, in contrast, demonstrated an increase in complexity over time. These changes were evident regardless of HIV coinfection status. These results may initially appear to conflict with a prior cross-sectional study of coinfecting and monoinfecting patients that determined greater HVR1 sequence variability in coinfecting patients, suggesting a possible accumulation of variants due to inability of those coinfecting to clear the dominant type.²⁹ However, while the duration of infection was not reported in that study, no patients had yet progressed to ESLD. Coinfecting patients analyzed at a relatively early point of HCV infection would yield comparable findings.

We observed similar trends in quasispecies diversity within the HVR1, assessed via mean Kimura 2-parameter distances between early and late sets of clones for the matched progressors and nonprogressors. Progressors exhibited greater early-timepoint genetic diversity than nonprogressors ($p = 0.04$) but that diversity narrowed significantly over time in progressors ($p = 0.04$) and remained constant in nonprogressors regardless of infection status. This is consistent with results from Tagariello et al.⁴² who recently analyzed samples from 12 HCV-monoinfecting hemophiliacs and found that those with

progressive disease had high homogeneity as assessed by heteroduplex mobility assay while nonprogressors had fairly static heterogeneity over time.

These differences within the HVR1 were confirmed in analysis of dN/dS ratios over time. After eliminating an outlying matched pair with a subtype shift in the progressor group, the progressors had slightly greater average dN/dS ratio at baseline than nonprogressors ($p = 0.25$). At the endpoint, however, the progressor group had a decreased average dN/dS ratio compared to no change for nonprogressors. Although the difference did not achieve statistical significance ($p = 0.31$), the trend is consistent with other measures of quasispecies evolution described above. Moreover, site-by-site analysis of predicted amino acid sequences for both early and late timepoints revealed significant differences in conservation of amino acid residues and, consequently, differences in hydrophobic properties between progressors and nonprogressors. While we have not yet determined specific structural and/or functional significance of these changes, the differences suggest that nonprogressors are able to maintain an equilibrium of quasispecies variants, either through immune selection or through clearance of emerging mutants. Progressors may have an abrogated ability to clear new variants, resulting in early complexity and diversity but the eventual dominance of a variant with greater replicative fitness or greater pathogenic potential.

Because the HCV core protein coding region has been associated with liver injury in vitro and in transgenic mouse models, we analyzed the same parameters of evolution over time in this portion of the HCV genome. Progressors demonstrated a slight increase in complexity over time compared to nonprogressors, although this change did not reach statistical significance. Likewise, although the early dN/dS ratio was significantly lower

in progressors than in nonprogressors, both groups of patients had dN/dS ratios consistently <1, indicating a lack of selective pressure for this region. In contrast to the findings in the HVR1, genetic distances in the core region between early and late clonal sets were likely to be greater for progressors ($p = 0.02$), indicating greater divergence. The core region has been shown to regulate oncogenes,^{11,15,52} alter cytokine expression,¹² and inhibit apoptosis,¹⁴ all of which are pathogenic mechanisms. Ray et al.⁵³ have postulated that specific mutations at core amino acid positions 9 and 11 help modulate NF- κ B activity. Our sequences displayed the RKT motif (positions 9-11) with four exceptions: T₁₁ \rightarrow P occurred in 3 clones, K₁₀ \rightarrow I occurred once, T₁₁ \rightarrow A once, and K₁₀ \rightarrow A in 11 clones. All of these mutations occurred in progressors, and all generate increased hydrophobicity at those positions. This suggests that increased rates of specific mutation could lead to higher rates of progression to fibrosis, possibly through inhibition of inflammatory repressor activity, although the exact mechanism of this process remains unknown. Analysis of CD4+/CD8+ proliferative responses to HCV-specific epitopes may help elucidate the factors associated with clearance of specific epitopes.

In conclusion, we have demonstrated numerous and significant differences in HCV quasispecies complexity and diversity in the HVR1 as well as differences in genetic diversity in the core region among interferon-naïve progressors and nonprogressors with inherited bleeding disorders. None of these differences were associated with HCV or HIV viral loads, demographic factors, or CD4+/CD8+ counts. Because patients infected through contaminated clotting factors were likely to have similar chances of exposure to multiple viral genotypes, subtypes, and variants through the large donor pool, these results suggest that multiple mechanisms, including viral fitness and host immune factors,

drive quasispecies emergence and selection in patients with progressive vs. nonprogressive disease. It is possible that more extensive cloning at each timepoint, and at interim timepoints, might clarify which mechanisms are operative during emergence and selection. Changes occurring in the HVR1 vs. the core region in progressors vs. nonprogressors are inconsistent. The HVR1 appears to exhibit the result of mutational selection; generally, the greater the immune selection process, the more variability will emerge over time. Patients progressing to ESLD, who likely exert less immune pressure on the virus, will demonstrate decreased variability in the HVR1 over time. The core region, however, undergoes stochastic rather than selectional mutation altering its hydrophobic character. In progressors, the changes that occur in the core region appear to be associated more frequently with hydrophobic elements that may have important functional significance. Further understanding of these mechanisms may help identify factors related to rate of progression and to treatment response in HCV-infected patients with inherited bleeding disorders.

Acknowledgements

We would like to thank Angelos Hatzakis, MD, PhD (Athens University Medical School, Athens, Greece) for critical review of this manuscript.

References

1. Aach RD, Szmuness W, Mosley JW, Hollinger FB, Kahn RA, Stevens CE, Edwards VM, Werch J. Serum alanine aminotransferase of donors in relation to the risk of non-A,non-B hepatitis in recipients: the transfusion-transmitted viruses study. *N Engl J Med*. 1981;304:989-994
2. Bica I, McGovern B, Dhar R, Stone D, McGowan K, Scheib R, Snyderman DR. Increasing mortality due to end-stage liver disease in patients with human immunodeficiency virus infection. *Clin Infect Dis*. 2001;32:492-497
3. Eyster ME, Diamondstone LS, Lien JM, Ehmann WC, Quan S, Goedert JJ. Natural history of hepatitis C virus infection in multitransfused hemophiliacs: effect of coinfection with human immunodeficiency virus. The Multicenter Hemophilia Cohort Study. *J Acquir Immune Defic Syndr*. 1993;6:602-610
4. Goedert JJ, Eyster ME, Lederman MM, Mandalaki T, De Moerloose P, White GC, 2nd, Angiolillo AL, Luban NL, Sherman KE, Manco-Johnson M, Preiss L, Leissinger C, Kessler CM, Cohen AR, DiMichele D, Hilgartner MW, Aledort LM, Kroner BL, Rosenberg PS, Hatzakis A. End-stage liver disease in persons with hemophilia and transfusion-associated infections. *Blood*. 2002;100:1584-1589
5. Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, Perelson AS. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science*. 1998;282:103-107
6. Farci P, Shimoda A, Coiana A, Diaz G, Peddis G, Melpolder JC, Strazzera A, Chien DY, Munoz SJ, Balestrieri A, Purcell RH, Alter HJ. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science*. 2000;288:339-344
7. Pawlotsky JM, Germanidis G, Frainais PO, Bouvier M, Soulier A, Pellerin M, Dhumeaux D. Evolution of the hepatitis C virus second envelope protein hypervariable region in chronically infected patients receiving alpha interferon therapy. *J Virol*. 1999;73:6490-6499
8. Farci P, Strazzera R, Alter HJ, Farci S, Degioannis D, Coiana A, Peddis G, Usai F, Serra G, Chessa L, Diaz G, Balestrieri A, Purcell RH. Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. *Proc Natl Acad Sci U S A*. 2002;99:3081-3086
9. Lyra AC, Fan X, Lang DM, Yusim K, Ramrakhiani S, Brunt EM, Korber B, Perelson AS, Di Bisceglie AM. Evolution of hepatitis C viral quasispecies after liver transplantation. *Gastroenterology*. 2002;123:1485-1493
10. Fan X, Di Bisceglie AM. Diversification of hypervariable region 1 of hepatitis C virus after liver transplantation. *J Med Virol*. 2003;70:212-218
11. Ray RB, Steele R, Meyer K, Ray R. Transcriptional repression of p53 promoter by hepatitis C virus core protein. *J Biol Chem*. 1997;272:10983-10986
12. Tsutsumi T, Suzuki T, Moriya K, Yotsuyanagi H, Shintani Y, Fujie H, Matsuura Y, Kimura S, Koike K, Miyamura T. Alteration of intrahepatic cytokine expression and AP-1 activation in transgenic mice expressing hepatitis C virus core protein. *Virology*. 2002;304:415-424
13. Tsutsumi T, Suzuki T, Moriya K, Shintani Y, Fujie H, Miyoshi H, Matsuura Y, Koike K, Miyamura T. Hepatitis C virus core protein activates ERK and p38 MAPK in cooperation with ethanol in transgenic mice. *Hepatology*. 2003;38:820-828

14. Ray RB, Meyer K, Ray R. Suppression of apoptotic cell death by hepatitis C virus core protein. *Virology*. 1996;226:176-182
15. Moriya K, Fujie H, Shintani Y, Yotsuyanagi H, Tsutsumi T, Ishibashi K, Matsuura Y, Kimura S, Miyamura T, Koike K. The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. *Nat Med*. 1998;4:1065-1067
16. Srinivas RV, Ray RB, Meyer K, Ray R. Hepatitis C virus core protein inhibits human immunodeficiency virus type 1 replication. *Virus Res*. 1996;45:87-92
17. Shih CM, Lo SJ, Miyamura T, Chen SY, Lee YH. Suppression of hepatitis B virus expression and replication by hepatitis C virus core protein in HuH-7 cells. *J Virol*. 1993;67:5823-5832
18. Sherman KE, Rouster SD, Mendenhall C, Thee D. Hepatitis cRNA quasispecies complexity in patients with alcoholic liver disease. *Hepatology*. 1999;30:265-270
19. Choo QL, Weiner AJ, Overby LR, Kuo G, Houghton M, Bradley DW. Hepatitis C virus: the major causative agent of viral non-A, non-B hepatitis. *Br Med Bull*. 1990;46:423-441
20. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*. 1989;244:359-362
21. Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*. 2001;17:1244-1245
22. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 1986;3:418-426
23. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4:406-425
24. Black SD, Mould DR. Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Anal Biochem*. 1991;193:72-82
25. Rosenthal E, Poiree M, Pradier C, Perronne C, Salmon-Ceron D, Geffray L, Myers RP, Morlat P, Pialoux G, Pol S, Cacoub P. Mortality due to hepatitis C-related liver disease in HIV-infected patients in France (Mortavic 2001 study). *Aids*. 2003;17:1803-1809
26. Jain MK, Skiest DJ, Cloud JW, Jain CL, Burns D, Berggren RE. Changes in mortality related to human immunodeficiency virus infection: comparative analysis of inpatient deaths in 1995 and in 1999-2000. *Clin Infect Dis*. 2003;36:1030-1038
27. Jaggy C, von Overbeck J, Ledergerber B, Schwarz C, Egger M, Rickenbach M, Furrer HJ, Telenti A, Battegay M, Flepp M, Vernazza P, Bernasconi E, Hirschel B. Mortality in the Swiss HIV Cohort Study (SHCS) and the Swiss general population. *Lancet*. 2003;362:877-878
28. Toyoda H, Fukuda Y, Koyama Y, Takamatsu J, Saito H, Hayakawa T. Effect of immunosuppression on composition of quasispecies population of hepatitis C virus in patients with chronic hepatitis C coinfecting with human immunodeficiency virus. *J Hepatol*. 1997;26:975-982

29. Sherman KE, Andreatta C, O'Brien J, Gutierrez A, Harris R. Hepatitis C in human immunodeficiency virus-coinfected patients: increased variability in the hypervariable envelope coding domain. *Hepatology*. 1996;23:688-694
30. Roque-Afonso AM, Robain M, Simoneau D, Rodriguez-Mathieu P, Gigou M, Meyer L, Dussaix E. Influence of CD4 cell counts on the genetic heterogeneity of hepatitis C virus in patients coinfecting with human immunodeficiency virus. *J Infect Dis*. 2002;185:728-733
31. Gaudy C, Moreau A, Veillon P, Temoin S, Lunel F, Goudeau A. Significance of pretreatment analysis of hepatitis C virus genotype 1b hypervariable region 1 sequences to predict antiviral outcome. *J Clin Microbiol*. 2003;41:3615-3622
32. Gerotto M, Sullivan DG, Polyak SJ, Chemello L, Cavalletto L, Pontisso P, Alberti A, Gretch DR. Effect of Retreatment with Interferon Alone or Interferon plus Ribavirin on Hepatitis C Virus Quasispecies Diversification in Nonresponder Patients with Chronic Hepatitis C. *J. Virol*. 1999;73:7241-7247
33. Noursbäum J-B, Polyak SJ, Ray SC, Sullivan DG, Larson AM, Carithers RL, Jr., Gretch DR. Prospective Characterization of Full-Length Hepatitis C Virus NS5A Quasispecies during Induction and Combination Antiviral Therapy. *J. Virol*. 2000;74:9028-9038
34. Koizumi K, Enomoto N, Kurosaki M, Murakami T, Izumi N, Marumo F, Sato C. Diversity of quasispecies in various disease stages of chronic hepatitis C virus infection and its significance in interferon treatment. *Hepatology*. 1995;22:30-35
35. Polyak SJ, McArdle S, Liu S-L, Sullivan DG, Chung M, Hofgartner WT, Carithers RL, Jr., McMahon BJ, Mullins JI, Corey L, Gretch DR. Evolution of Hepatitis C Virus Quasispecies in Hypervariable Region 1 and the Putative Interferon Sensitivity-Determining Region during Interferon Therapy and Natural Infection. *J. Virol*. 1998;72:4288-4296
36. Sarrazin C, Bruckner M, Herrmann E, Ruster B, Bruch K, Roth WK, Zeuzem S. Quasispecies heterogeneity of the carboxy-terminal part of the E2 gene including the PePHD and sensitivity of hepatitis C virus 1b isolates to antiviral therapy. *Virology*. 2001;289:150-163
37. Garcia JM, Serrano PL, Terron Sdel C, Marugan RB, Garcia GM, Grande LG, Miquel J, Plaza AG. Very rapid evolution of infection with hepatitis C virus transmitted by an accidental needlestick. *Clin Infect Dis*. 2003;36:1632-1634
38. Hughes MG, Jr., Rudy CK, Chong TW, Smith RL, Evans HL, Iezzoni JC, Sawyer RG, Pruett TL. E2 quasispecies specificity of hepatitis C virus association with allografts immediately after liver transplantation. *Liver Transpl*. 2004;10:208-216
39. Silini E, Belli L, Brambilla S, Foti L, Gritti C, Lisa A, Alberti AB, Vinci M, De Carlis L, Rondinara G, Pinzello G. Sequence variation in the hypervariable region 1 of hepatitis C virus and posttransplantation recurrent hepatitis. *Liver Transpl*. 2003;9:1040-1047
40. Sullivan DG, Wilson JJ, Carithers RL, Jr., Perkins JD, Gretch DR. Multigene Tracking of Hepatitis C Virus Quasispecies after Liver Transplantation: Correlation of Genetic Diversification in the Envelope Region with Asymptomatic or Mild Disease Patterns. *J Virol*. 1998;72:10036-10043
41. Takahashi K, Takahashi T, Takahashi S, Watanabe K, Boku S, Matsui S, Arai F, Asakura H. Difference in quasispecies of the hypervariable region 1 of hepatitis C

virus between alcoholic and non-alcoholic patients. *J Gastroenterol Hepatol*. 2001;16:416-423

42. Tagariello G, Gerotto M, Pontisso P, Belvini D, Salviato R, Radossi P, Alberti A. Hepatitis C virus quasispecies in the natural course of HCV-related disease in patients with haemophilia. *Haemophilia*. 2004;10:81-86

43. Pawlotsky JM, Pellerin M, Bouvier M, Roudot-Thoraval F, Germanidis G, Bastie A, Darthuy F, Remire J, Soussy CJ, Dhumeaux D. Genetic complexity of the hypervariable region 1 (HVR1) of hepatitis C virus (HCV): influence on the characteristics of the infection and responses to interferon alfa therapy in patients with chronic hepatitis C. *J Med Virol*. 1998;54:256-264

44. Toyoda H, Kumada T, Nakano S, Takeda I, Sugiyama K, Osada T, Kiriyaama S, Sone Y, Kinoshita M, Hadama T. Quasispecies nature of hepatitis C virus and response to alpha interferon: significance as a predictor of direct response to interferon. *J Hepatol*. 1997;26:6-13

45. Neau D, Jouvencel AC, Legrand E, Trimoulet P, Galperine T, Chitty I, Ventura M, Le Bail B, Morlat P, Lacut JY, Ragnaud JM, Dupon M, Fleury H, Lafon ME. Hepatitis C virus genetic variability in 52 human immunodeficiency virus-coinfected patients. *J Med Virol*. 2003;71:41-48

46. Cao K, Mizokami M, Orito E, Ohno T, Nakano T, Ding X, Ohba K, Mizuno Y, Hayashi K, Kumada K. Nucleotide and amino acid diversities of E2/NS1 hypervariable region 1 and response to interferon therapy in patients with chronic hepatitis C virus infection. *Hepatology Research*. 1998;11:166-177

47. Karada K, Watanabe H, Shijo H, Okumura M. Predictive factors regarding the responsiveness to interferon therapy for patients with chronic hepatitis C: an assessment of the hypervariable region 1 quasispecies in genotype 1b of chronic hepatitis C. *International Hepatology Communications*. 1996;5:215-220

48. Curran R, Jameson CL, Craggs JK, Grabowska AM, Thomson BJ, Robins A, Irving WL, Ball JK. Evolutionary trends of the first hypervariable region of the hepatitis C virus E2 protein in individuals with differing liver disease severity. *J Gen Virol*. 2002;83:11-23

49. Leone F, Zylberberg H, Squadrito G, Le Guen B, Berthelot P, Pol S, Brechot C. Hepatitis C virus (HCV) hypervariable region 1 complexity does not correlate with severity of liver disease, HCV type, viral load, or duration of infection. *J Hepatol*. 1998;29:689-694

50. Harris KA, Teo CG. Diversity of hepatitis C virus quasispecies evaluated by denaturing gradient gel electrophoresis. *Clin Diagn Lab Immunol*. 2001;8:62-73

51. Devereux H, Telfer P, Brown D, Morris A, Dusheiko G, Emery V, Lee C. Longitudinal genotype analysis and quantification of hepatitis C virus in haemophilic patients receiving interferon-alpha therapy. *J Viral Hepat*. 1996;3:43-48

52. Ray RB, Lagging LM, Meyer K, Ray R. Hepatitis C virus core protein cooperates with ras and transforms primary rat embryo fibroblasts to tumorigenic phenotype. *J Virol*. 1996;70:4438-4443

53. Ray RB, Steele R, Basu A, Meyer K, Majumder M, Ghosh AK, Ray R. Distinct functional role of Hepatitis C virus core protein on NF-kappa B regulation is linked to genomic variation. *Virus Res*. 2002;87:21-29

54. Eyster ME, Sherman KE, Goedert JJ, Katsoulidou A, Hatzakis A. Prevalence and changes in hepatitis C virus genotypes among multitransfused persons with hemophilia. The Multicenter Hemophilia Cohort Study. J Infect Dis. 1999;179:1062-1069

Tables

Table 1. Demographic characteristics of progressors to end-stage liver disease (cases) and nonprogressors (controls).

	Progressors (n = 10)	Nonprogressors (n = 10)	p-value
Race (n, %)			
Caucasian	10 (100)	6 (60)	n.s.
Other	--	4 (40)	
HIV-positive [n (%)]	8 (80)	8 (80)	n.s.
HCV viral load (bDNA, copies/mL) [mean (range)]	2.27E+07 (5.19e+05 – 6.88e+07)	2.38E+07 (5.55e+05 – 8.96e+07)	n.s.
HCV subtypes*	1 = 1 1a = 4 1b = 5	1 = 1 1a = 7 1b = 2	n.s.
Baseline CD4+ (cells/mL) [mean (range)]	525 (249 – 1388)	417 (212 – 680)	n.s.
Baseline CD8+ (cells/mL) [mean (range)]	698 (277 – 1379)	922 (133 – 2592)	n.s.
Time from baseline to follow-up (years) [mean (range)]	7.82 (4.7 – 13.9)	10.71 (6.1 – 14.3)	0.01
Baseline AST (U/L) [mean (range)]	70 (38 – 128)	47.6 (29 – 64)	n.s.
Baseline ALT (U/L) [mean (range)]	90 (48 – 130)	87 (34 – 197)	n.s.

* 3 patients exhibited genotype shifts during the study period: one from 1a to 1b, one from 1b to 1a, and one from 1b to 1.⁵⁴

Table 2. Mean frequencies of unique HVR1 nucleotide sequences for each patient subpopulation.
Frequencies were determined by the proportion of unique nucleotide sequences to the number of clones assessed for each patient at early and late timepoints. “N” represents number of patients in each population.

Timepoint	Mean complexity (standard error)		
	Progressors (n = 10)	Nonprogressors (n = 10)	Difference: Progressor - Nonprogressor
Coinfected (n=16)			
Early	0.78 (SE 0.02)	0.54 (SE 0.09)	0.24
Late	0.65 (SE 0.07)	0.72 (SE 0.07)	-0.07
Monoinfected (n=4)			
Early	0.53 (SE 0.03)	0.81 (SE 0.18)	-0.28
Late	0.48 (SE 0.23)	1.0 (SE 0.0)	-0.52
Total			
Early	0.73 (SE 0.04)	0.59 (SE 0.08)	0.14
Late	0.61 (SE 0.04)	0.77 (SE 0.07)	-0.16
Late - Early	-0.12 (SE 0.07)	0.18 (SE 0.10)	-0.30*

***P = 0.01** for difference in unique nucleotide sequence frequencies between progressors and nonprogressors over time (generalized estimating equation model with binomial distribution and logit link)

Table 3. Mean dN/dS ratio comparisons within the HVR1 for progressors (cases) and nonprogressors (controls). The numbers of synonymous mutations (dS) and the numbers of nonsynonymous mutations (dN) were determined using the Nei-Gojobori method with the Jukes-Cantor correction for multiple substitutions.²² The mean dN/dS ratio, a measure of selective pressure, was calculated for each patient. One matched pair is excluded due to a shift in dominant subtype in the progressor. Progressors demonstrated an overall decrease in dN/dS ratio over time while nonprogressors remain stable.

Timepoint	dN/dS ratio (standard error)		
	Progressors	Nonprogressors	p-value for difference between progressors and nonprogressors
Early dN/dS ratio	2.02 (SE 0.70)	0.98 (SE 0.22)	p = 0.25
Late dN/dS ratio	0.62 (SE 0.22)	0.98 (SE 0.39)	p = 0.52
Mean difference (early to late) in dN/dS ratio	- 1.40 (SE 0.83)	- 0.006 (SE 0.55)	p = 0.31
p-value for difference over time	p = 0.13	p = 0.99	

Table 4. Mean frequencies of unique core region nucleotide sequences for each patient subpopulation. Frequencies were determined by the proportion of unique nucleotide sequences to the number of clones assessed for each patient at early and late timepoints within the core coding region. “N” represents number of patients in each population.

Timepoint	Mean complexity (standard error)		
	Progressors (n = 10)	Nonprogressors (n = 10)	Difference: Progressor - Nonprogressor
Coinfected (n=16) Early	0.56 (SE 0.11)	0.72 (SE 0.06)	-0.16
Late	0.61 (SE 0.10)	0.73 (SE 0.07)	-0.12
Monoinfected (n=4) Early	0.47 (SE 0.20)	0.92 (SE 0.08)	-0.45
Late	0.89 (SE 0.11)	0.83 (SE 0.08)	0.06
Total Early	0.54 (SE 0.09)	0.76 (SE 0.05)	-0.22
Late	0.67 (0.09)	0.75 (SE 0.06)	-0.08
Late - Early	0.13 (SE 0.13)	-0.01 (SE 0.06)	0.14

* P = n.s. for all comparisons

Table 5. Mean dN/dS ratios within the core coding region for early and late timepoints for progressors (cases) vs. nonprogressors (controls). The numbers of synonymous mutations (dS) and the numbers of nonsynonymous mutations (dN) were determined using the Nei-Gojobori method with the Jukes-Cantor correction for multiple substitutions.²² The mean dN/dS ratio, a measure of selective pressure, was calculated for each patient.

Timepoint	dN/dS ratio (standard error)		
	Progressors (n=10)	Nonprogressors (n=10)	p-value for difference between progressors and nonprogressors
Early dN/dS ratio	0.15 (SE 0.07)	0.30 (SE 0.10)	p = 0.05
Late dN/dS ratio	0.10 (SE 0.03)	0.16 (SE 0.04)	p = 0.26
Mean difference (early to late) in dN/dS ratio	-0.05 (SE 0.08)	-0.13 (SE 0.10)	p = 0.31
p-value for difference over time	p = 0.54	p = 0.21	

Figures

Figure 1. Mean genetic distances within the HVR1 for early and late clonal sets from progressors and nonprogressors. Distance was calculated via the mean Kimura 2-parameter pairwise distances for each patient.

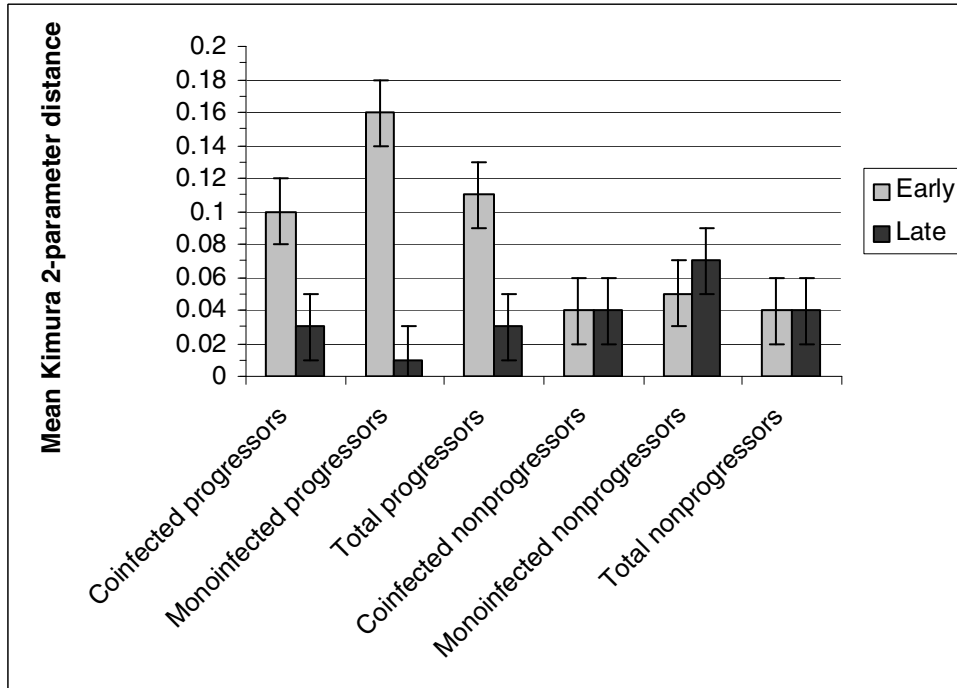
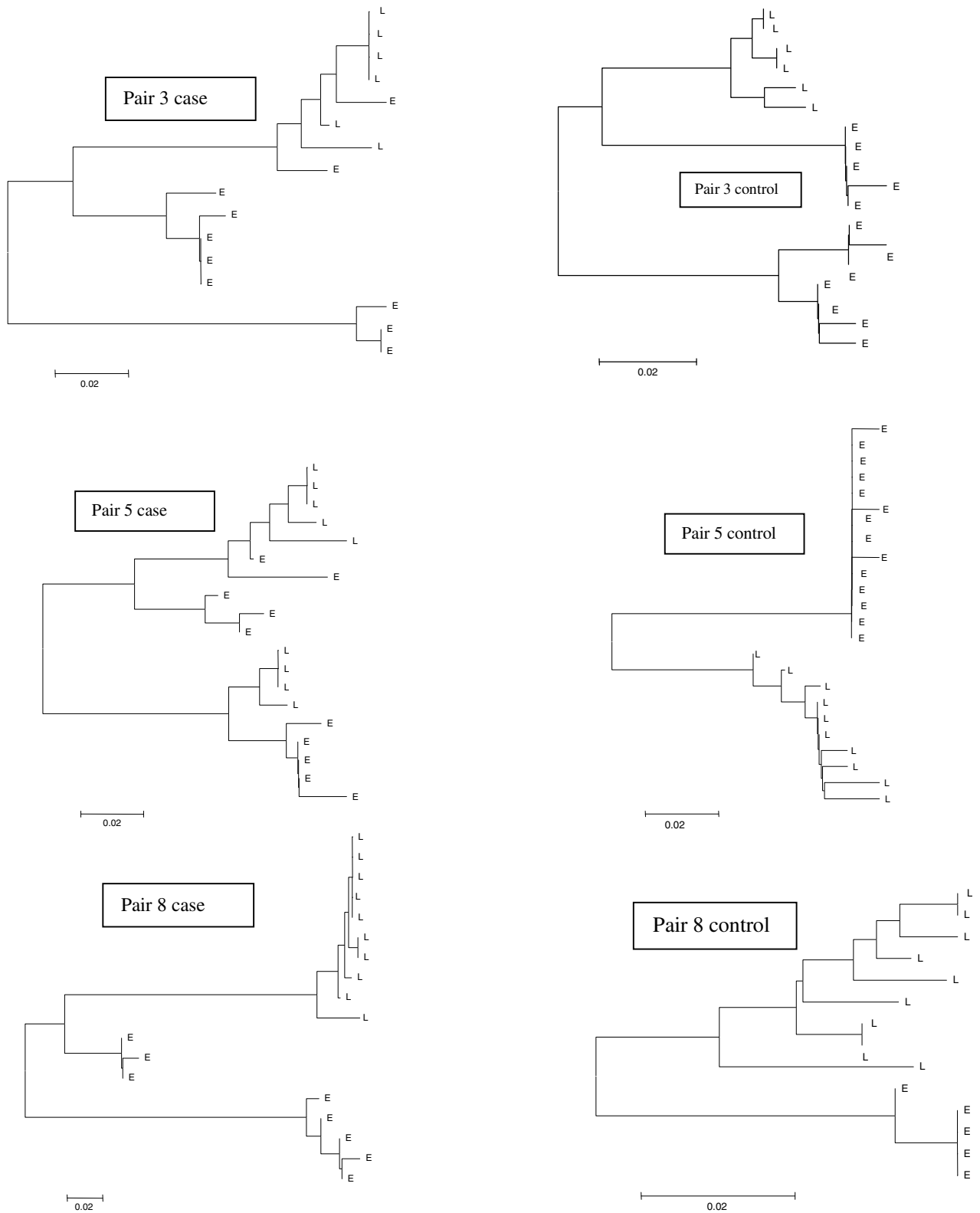
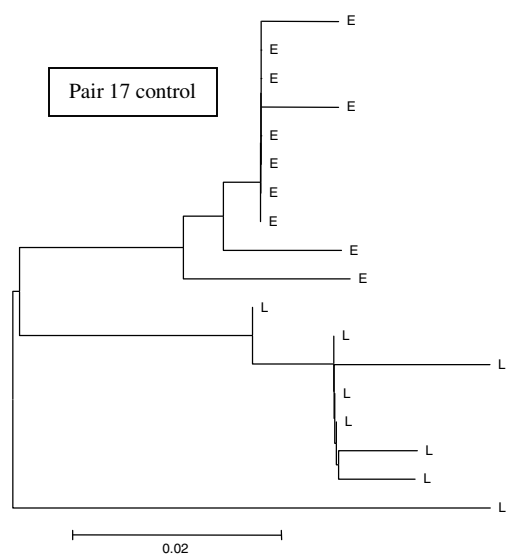
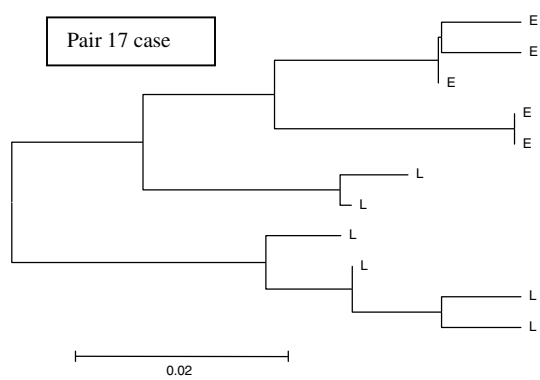
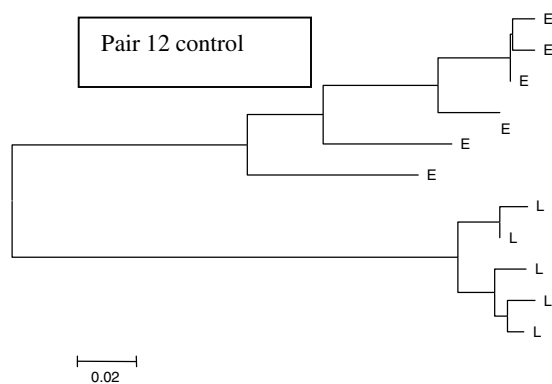
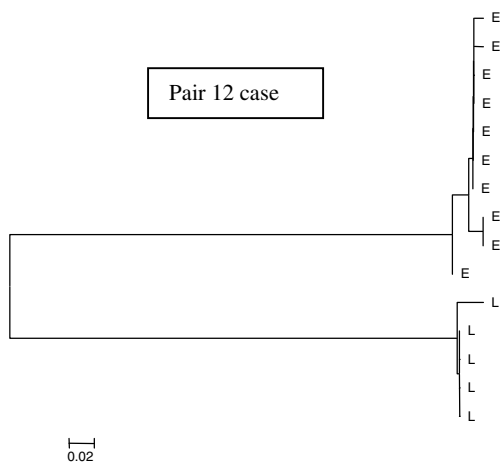
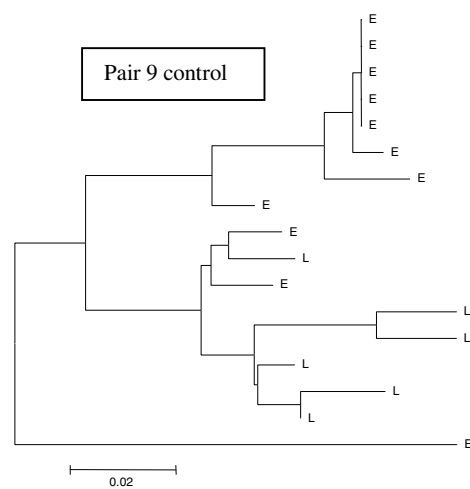
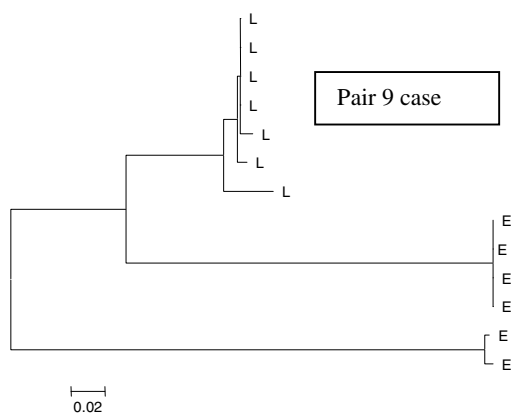


Figure 2. Phylogenetic trees (HVR1) for cases and controls.





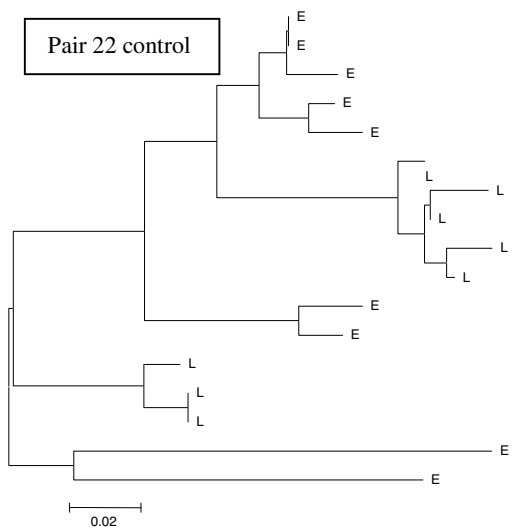
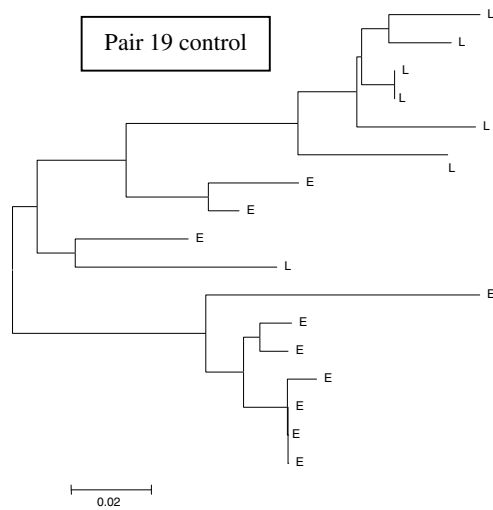
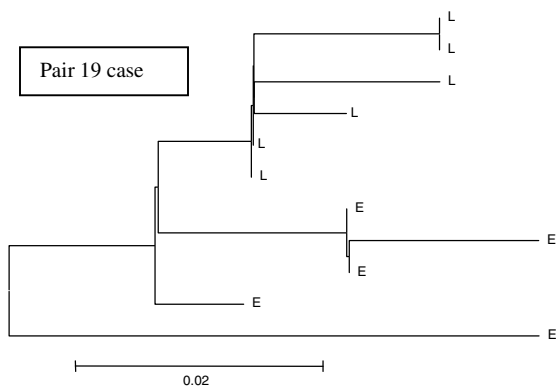
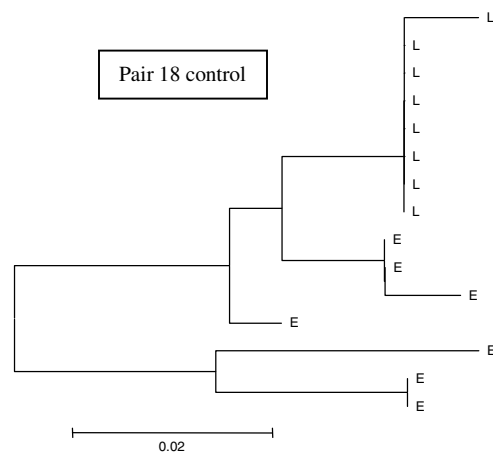
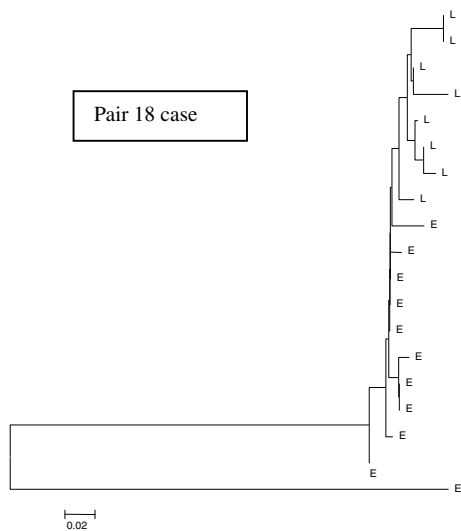


Figure 3. Predicted HVR1 amino acid sequences.

Pair 2 case early
 #4208 QGTY-N-E-- VSP-PPP-V- --RACSHL-P SQ
 #4203A-
 #4204
 #4205 .S.....
 #4209H.....
 #4210D.....R.....
 #4211 .S.....S.....
 #4212
 #4213
 #4216
 #4218 ...SG.....
 #4220
 #4221L.....
 #4222
 #4223G.....
 #4224

Pair 3 case early
 #3I48-02 ETTVTGAVAG RTTAVFAGLF RSGAKQD
 #3I48-03
 #3I48-04 ...I..G... .DA.G.V.F. AR..Q..
 #3I48-05
 #3I48-06 ...I..G... .DA.G.V.F. AR..Q..
 #3I48-07GTVSRL.SI. S.....
 #3I48-08 ...I..G... .DA.G.V.F. AR..Q.N
 #3I48-09GTVSRL.... S.....
 #3I48-10
 #3I48-11I..P.....

Pair 5 case early
 #5I10-01 GTYVVGGTAA QSAYGLTSL L APGAKQN
 #5I10-02 D...T.AITG .VS..A..F T...R..
 #5I10-03 D...T.AITG .GVS..A..F N...R..
 #5I10-05 ..H.....
 #5I10-06 N...T..STG RG.S.IAG.F V.....
 #5I10-09 N...T.ASTG RG.S.IVG.F T...R..
 #5I10-10 ..H.....
 #5I10-11 ..H.....
 #5I10-14 D...T.AITG .GVS..A..F N...R..
 #5I10-15 ..H.....

Pair 8 case early
 #8I69-05 ETHVTGGTVA HATAGFTRLL APGAKQN
 #8I69-08 ...TV..K.. Q.A...V.F. E...R..
 #8I69-09 ...TV..K.. Q.A...V.F. E...R..
 #8I69-11 ...TV..K.. Q.A...V.F. E...R..
 #8I69-13 ...TV..K.. Q.A...V.F. E...R..
 #8I69-14
 #8I69-15 ...TV..K.. Q.A...V.F. E...R..
 #8I69-16

Pair 9 case early
 #9I73-03 QGTYATGGAS ARTTASFAGL FTLGPSQK
 #9I73-06 E-.HVS..SA GCAALGLT.. .SP.AK.N
 #9I73-09 .S.....
 #9I73-11 E-.HVS..SA G.AALGLT.. .SP.AK.N
 #9I73-12
 #9I73-14 .S.....

Pair 2 case late
 #4403 -TH-H---- -R-AP--AWT YW-SSVRAP- QNIQLI
 #4404 E...-S--E- -...-...-...-...-...
 #4405 -...-...-...-...-...-...
 #4406 -...-...-...-...-...-...
 #4407 -...-...-...-...-...-...
 #4408 E...-L-...-...-...-...-...
 #4409 -...-...-...-...-...-...Q..
 #4411 -...-...-...-...-...-...
 #4412 E...-L-CE-- RP--ALGL-- -...-...-...

Pair 3 case late
 #3I50-04 ETTVIGGIVS RITARLASIF SPGAKQD
 #3I50-05S.....
 #3I50-06D..
 #3I50-07S.....
 #3I50-08S.....
 #3I50-10S.....

Pair 5 case late
 #5I9-02 NTYVTGGSTG RGAAGIAGLF VPGAKQN
 #5I9-04 E.H.V...AA QS.F.L.S.L A.....
 #5I9-07 D.....
 #5I9-08
 #5I9-10 E.H.V...AA QS.F.L.S.L A.....
 #5I9-11
 #5I9-13 E.P.V...AA QS.F.LTS.L A.....
 #5I9-14
 #5I9-15 E.H.V...AA QS.F.L.S.L A.....

Pair 8 case late
 #8I71-02 ETYVSGGSVG RTVAGVGSLF NPGAKQN
 #8I71-03
 #8I71-04
 #8I71-07
 #8I71-09 K.....A.. .I.....
 #8I71-11
 #8I71-12
 #8I71-14
 #8I71-15
 #8I71-19

Pair 9 case late
 #9I74-03 NIHTTGGTAA YNARGLTSLF APGPSQR
 #9I74-04
 #9I74-05
 #9I74-08
 #9I74-10C...I. T.....
 #9I74-11
 #9I74-12

Pair 2 control early
 #2C38-06 ETHITGGSAG HATSGFVRL L APGAKQN
 #2C38-08
 #2C38-11
 #2C38-12 ..R.....
 #2C38-14
 #2C38-15
 #2C38-16Q..
 #2C38-17
 #2C38-20

Pair 3 control early
 #3C45-01 GTHVTGGREA SAAHGLTKLL TVGPKQN
 #3C45-02
 #3C45-03 Q...S..S.. R..YK.SSFF ...A..
 #3C45-04 Q.....S.. R..YK.SSFF .L.A..
 #3C45-05 Q.....S.. R..YK.SSFF .L.A..
 #3C45-06 Q...S..S.. R..YK.SSFF ...A..
 #3C45-07
 #3C45-08 Q...S..S.. R..YK.SSFF ...A..
 #3C45-09
 #3C45-10
 #3C45-11 Q...S..S.. R..YK.SSFF ...A..
 #3C45-12 Q.....S.. R..YK.SSFF .L.A..
 #3C45-13

Pair 5 control early
 #5C1-02 QTRVTGGSAS HTTSGLVGLF RSGPKQN
 #5C1-03
 #5C1-04
 #5C1-07
 #5C1-08
 #5C1-10
 #5C1-11
 #5C1-13
 #5C1-14
 #5C1-15
 #5C1-17R.....
 #5C1-18
 #5C1-19
 #5C1-20

Pair 8 control early
 #8C66-03 TTYATGGAQG RTTQGLTSLF TPGSSQN
 #8C66-06
 #8C66-08
 #8C66-12
 #8C66-14

Pair 9 control early
 #9C76-01 GTYVTGGNVG HTTAGLAGLL APGARQN
 #9C76-02
 #9C76-03A.....
 #9C76-04
 #9C76-05 D.....I.... TK..K..
 #9C76-07 D.....A.....F. TK..K..
 #9C76-08 E.....S.. .NVK.....TA.PK..
 #9C76-09
 #9C76-10 E.....S.. .A.....
 #9C76-11
 #9C76-12

Pair 2 control late
 #2C41-04 ETHITGGSAG HATSGFVHLL APGAKQN
 #2C41-05 ...V.....SP.
 #2C41-06
 #2C41-07
 #2C41-09R..
 #2C41-10
 #2C41-11 ...V.....S..
 #2C41-12

Pair 3 control late
 #3C47-01 QTHVTGREA RAAQGFAQLF TAGAKQN
 #3C47-03S...
 #3C47-05S...
 #3C47-06S...
 #3C47-07T.. ...H.....
 #3C47-12S...

Pair 5 control late
 #5C6-01 ETQVTGGSAG YTTYGLAGLF QPGAKQN
 #5C6-02C.....
 #5C6-04A.....
 #5C6-06N.....
 #5C6-08
 #5C6-10H.....R.....
 #5C6-12
 #5C6-13
 #5C6-14
 #5C6-16R.....

Pair 8 control late
 #8C68-01 TTYTTGGAQG RTTQGLTSLF KPGSSQN
 #8C68-07S.....F....R.....
 #8C68-08 N..A.....S.....F....T.....
 #8C68-09 N..A.....S.....F....R.....
 #8C68-10S.....F....R.....
 #8C68-12S.....F....R.....
 #8C68-13 D.....S.....F....R.....
 #8C68-14 N..A.....S.....F....R.....
 #8C68-15S.....F....R.....

Pair 9 control late
 #9C75-02 NIYVIGMVG HTTAGLAGLL TKGARQN
 #9C75-03 D.....SA.K..
 #9C75-06
 #9C75-08
 #9C75-09A. R.....
 #9C75-10 ..C.....A.

Pair 12 case early

#12I16-02 TTYTTGGSAA LNVRGVVGLEF QRGARQD
 #12I16-03
 #12I16-04
 #12I16-07
 #12I16-16 A.....
 #12I16-18
 #12I16-21
 #12I16-22 F...I....
 #12I16-24 F...I....
 #12I16-26

Pair 17 case early

#17I117-02 RRSNGNPRHRG KCRPHHGWPCLSLHTRRQAE
 #17I117-06
 #17I117-07 ..R.....A...R.....
 #17I117-10 ..R.....A...W...R.....
 #17I117-11 ..R.....A...R.....

Pair 18 case early

#18I125-01 RTYTTGGQVA HGTSVFTSLF KRGPKQN
 #18I125-02Q.....
 #18I125-03Q...D
 #18I125-04 -.V..ESS. RA.HG.....SP.AS.K
 #18I125-05Q.....
 #18I125-06Q...D
 #18I125-07Q.....
 #18I125-08Q.....
 #18I125-09Q.....
 #18I125-10Q.....
 #18I125-12Q...D

Pair 19 case early

#19I131-02 GTHVTGGTLA RTTQGFSTFF ASGPSQK
 #19I131-03 ..R....AQGL.....
 #19I131-05 S.R....AQGL.....
 #19I131-08 ..R....AQGL.....
 #19I131-09 ..R....AQGL.....

Pair 22 case early

#22I143-02 --HVTGGSAG HATSGVC*PS CTRRQAE
 #22I143-15 --.....R.AL.LTGLF SPGAKQN
 #22I143-16 --.....FVSLLE APGAKQN
 #22I143-20 --.....R.AL.LTGLF SPGAKQN
 #22I143-21 --.....R.AL.LTGLF SPGAKQN
 #22I143-22 --YT...VA QGAF.LASFL SPGAKQN
 #22I143-23 --YT...VA QGAF.LANFL SRGPKQN
 #22I143-24 --.S.....R.AL.LTGLF SPGAKQN
 #22I143-25 --YT...VA QGAF.LANFL SRGPKQN
 #22I143-26 --.....R.AL.LTGLF SPGAKQN
 #22I143-27 --YT...VA QGAF.LASFL SPGAKQN
 #22I143-28 --F.S..T.. .TL..FTSLLE SPGAKQN
 #22I143-29 --.S...EA R.AYKLSSFF TVGAKQN
 #22I143-30 --.S.....R.AL.LTGLF SPGAKQN
 #22I143-31 STYA...ASA RT.ASFAGLE TLGPSQK

Pair 12 case late

#12I18-04 ETFVSGGTAG HTLS-FTSLLE SPGAKQN
 #12I18-08G.....
 #12I18-10G.....
 #12I18-16G.....
 #12I18-18G.....

Pair 17 case late

#17I120-01 ETHVTGGSAG RTTAGFASFF TQGAQKN
 #17I120-04H.....L. NP.....
 #17I120-06H.....L. NP.....
 #17I120-09 ..A.....H.....L.
 #17I120-11H.....
 #17I120-12

Pair 18 case late

#18I127-01 HTYTTGGQVA HGTSVLTSLE KQGPKN
 #18I127-02 R..A.....
 #18I127-04 R.....F.....S
 #18I127-05 R..A.....
 #18I127-06 S.....
 #18I127-07
 #18I127-09 R.....
 #18I127-11

Pair 19 case late

#19I134-03 GTRVTGGAQG RTTKGLTSFF VSGPSQK
 #19I134-04Q.....
 #19I134-07Q.....
 #19I134-09Q.....
 #19I134-16
 #19I134-17Q.....

Pair 22 case late

#22I146-01 ETYTTGGSVA QGAFGLASFL SPGAKQN
 #22I146-03
 #22I146-04
 #22I146-07
 #22I146-08
 #22I146-09
 #22I146-11S.....

Pair 12 control early

#12C13-01 TTHVTGGNAA FTTSTFVSIW NPGAKQN
 #12C13-02S.....KL.....
 #12C13-04S.....G...L.....
 #12C13-05S.....*L.....
 #12C13-06

Pair 17 control early

#17C121-01 HTTVSGGVQA YNTRGLTSLE SPGAQQK
 #17C121-02
 #17C121-04
 #17C121-05
 #17C121-06
 #17C121-07
 #17C121-08C.....
 #17C121-09N.*.....R
 #17C121-10
 #17C121-12

Pair 18 control early

#18C128-01 ETYVTGGVAG HAARGFAGLE SPGAKQN
 #18C128-02S.L...SI. DR.....
 #18C128-03S.L...SI. DR.....
 #18C128-04L...R. DR.....
 #18C128-07
 #18C128-10RS.L.L....Q.....
 #18C128-12S.L...SI. DR.....

Pair 19 control early

#19C136-03 NTHVTGGSVA SHLNGDPR-S LH-GARQN
 #19C136-08R.....-..-.....
 #19C136-10 Q.Y.....LRRG...Q-. S.-.K..
 #19C136-11 Q.Y.....LRRG..QQ-. S.-.....
 #19C136-12HDE.....-..-.....
 #19C136-13HD.....-..-.....
 #19C136-14 H.Y.....DHD..QQ-. S.-.....
 #19C136-16A...GRGSAV..S..K..
 #19C136-17-..-.....
 #19C136-19-..-.....

Pair 22 control early

#22C139-01 ETHVTGGSAA REANIFANIF SPGAKQN
 #22C139-02
 #22C139-03G .D.SV..G...S.....
 #22C139-04G .D.SL..G...S.....
 #22C139-05KSVYS..SFL QR.....
 #22C139-08G .D.SL..G...S.....
 #22C139-10G .D.SL..G...S.....
 #22C139-11G .D.SV..G...S.....
 #22C139-12T. .A.AGITGL. .L.....

Pair 12 control late

#12C3-02 TTQISGGRAA QVTFATSLLE NSGPKQN
 #12C3-03K.. EAA...S.FF QL..R..
 #12C3-04
 #12C3-05K.. .A.....F. SR.....
 #12C3-07
 #12C3-08K.. .GA...ANFF Q.....

Pair 17 control late

#17C124-02 HTTVLGGVQA YNTRGLTSLE SPGPQQK
 #17C124-03
 #17C124-04
 #17C124-05
 #17C124-06
 #17C124-08G.....
 #17C124-10
 #17C124-11 ..A.S..A.. .D.....T..AR..

Pair 18 control late

#18C130-02 ETYVTGGVAG HNALGFASLE DRGAKQN
 #18C130-04
 #18C130-06
 #18C130-07
 #18C130-08
 #18C130-09
 #18C130-10
 #18C130-11

Pair 19 control late

#19C138-01 QTHVTGGSAG HTALGIASLE SPGARQN
 #19C138-03 ..R.....R..S...N.F .Q.....
 #19C138-05 H....A...Q.VA..S..F TS.PK..
 #19C138-06S...N..K..
 #19C138-07 ..R.....Y.TS...N.FK..
 #19C138-08 H.....S...N..
 #19C138-10 H.....S...N..
 #19C138-11

Pair 22 control late

#22C142-01 ETHVTGGAAG RDAGLLTSIS SPGAKQN
 #22C142-02STA HG.LG.A.FL .S.....
 #22C142-03STA HG.LG.A.FL .S.....
 #22C142-04 G.....T...R.....F
 #22C142-05T...F
 #22C142-06STA HG.LG.A.LL .S.....
 #22C142-07 G.....T...F
 #22C142-11T...F

Figure 4. Differences in number of amino acids present at sites within the HVR1 for progressors and nonprogressors.

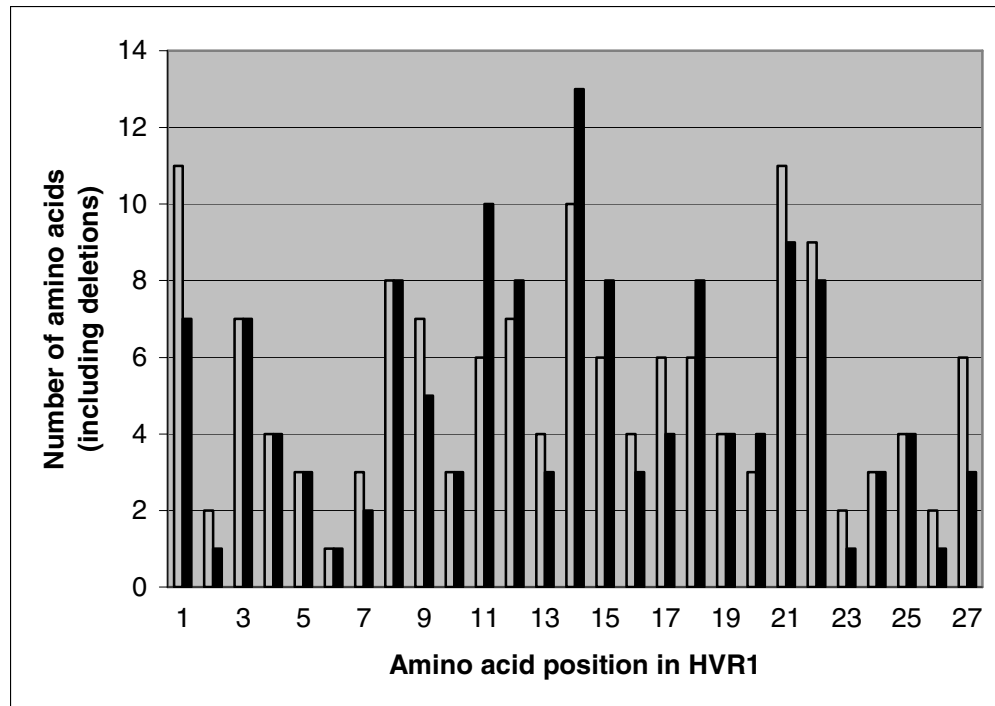


Figure 5. Hydropathic character of amino acid residues at variable sites within the HVR1 for progressors and nonprogressors.

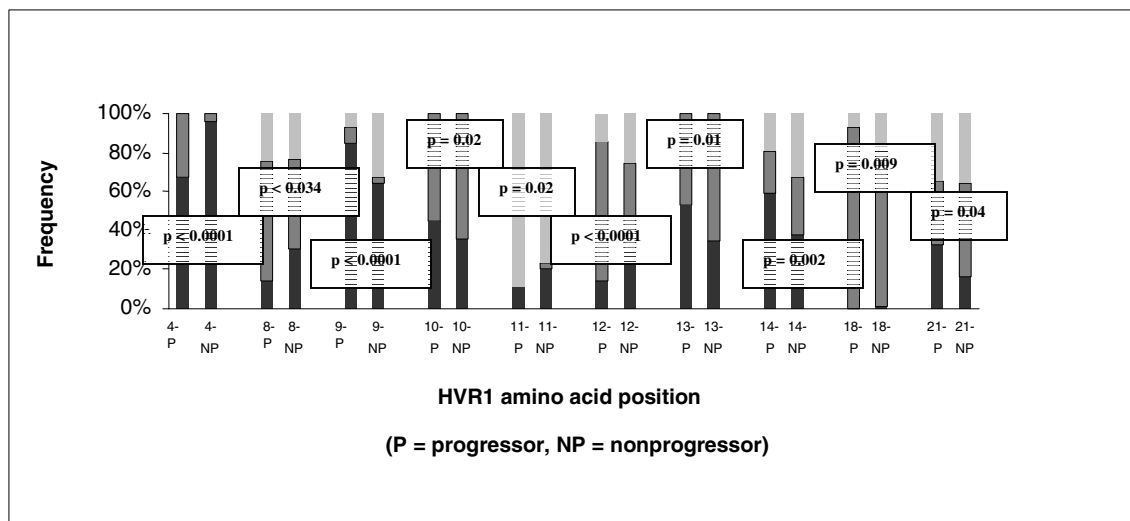


Figure 6. Mean genetic distances for core region within early and late clonal sets from progressors and nonprogressors. Distance was calculated via the mean Kimura 2-parameter pairwise distances for each patient.

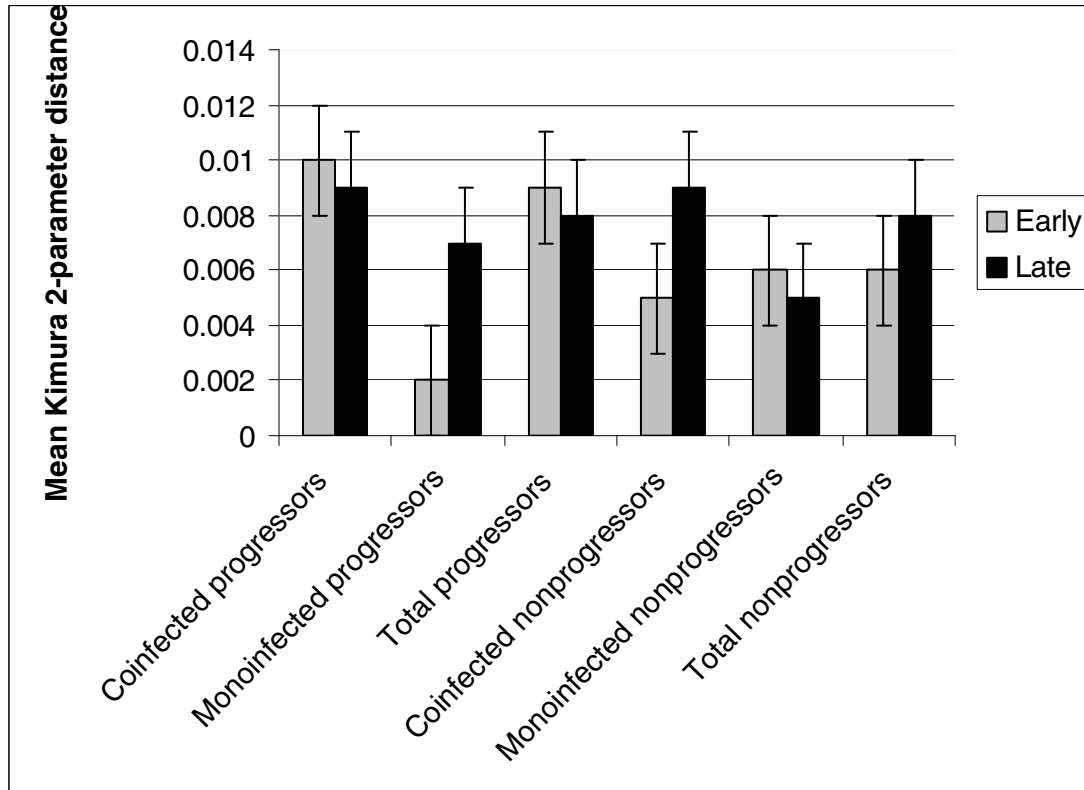


Figure legends

Figure 1. Mean genetic distances within the HVR1 between early and late clonal sets for progressors and nonprogressors. Distance was calculated via the mean Kimura 2-parameter pairwise distances for each patient. Regardless of HIV status, progressors as a group tended to have greater genetic diversity at the early timepoint than nonprogressors ($p = 0.04$) and demonstrated significant narrowing of genetic diversity over time ($p = 0.04$). Nonprogressors as a group maintained the same level of diversity over the study duration.

Figure 2. Phylogenetic trees from early/late E2/NS1 HVR1 clonal sets from ESLD progressors (cases) and nonprogressors (controls), constructed in MEGA2.1 using the neighbor-joining method of Saitou and Nei¹⁸. Sequences from early-timepoint clones are labeled with an “E” at the end of the sequence name. Late-timepoint sequences are labeled with an “L”. One matched pair is not represented here due to a subtype shift in the progressor that precluded tree generation in MEGA. Progressors trended toward greater genetic diversity at the early timepoint than nonprogressors ($p = 0.04$) and demonstrated significant narrowing of genetic diversity over time, regardless of infection status ($p = 0.04$). Nonprogressors maintained the same level of diversity over the study duration.

Figure 3. Predicted 81-bp HVR1 amino acid sequences for early and late clonal sets from progressors to end-stage liver disease (cases) and nonprogressors (controls). Sequence numbers are identified along the left column. Dots indicate no change from initial sequence (top line).

Figure 4. Number of amino acids, including deletions, represented at each site within the HVR1. Progressors (cases) are represented in gray, and nonprogressors (controls) in black. The mean number of amino acid substitutions per site, including deletions, was 5.3 (SE 0.54) for the progressors and 5 (SE 0.61) for nonprogressors ($p = 0.44$). Certain sites (2, 6, 7, 23, and 26) were highly conserved for both cases and controls but others (1, 9, 13, 27) had different numbers of substitutions between groups. Sites 4, 9, 13, and 27 had significantly less of the predominant amino acid residue in the case group than in the control group (p -values = 0.002, 0.03, 0.02, 0.01, and 0.02 respectively). For site 2, this was due to one subject's early set of clones with a deletion. For the other sites, this was due to substitutions among several sets of clones.

Figure 5. Amino acid sites from the HVR1 with statistically significant variability in hydrophobic character. Hydrophobicity was determined using the scale of Black and Mould.¹⁹ Hydrophobic residues include F, I, W, Y, L, V, M, P, C, and A. Neutral residues include G, S, and T. Hydrophilic residues include K, Q, N, H, E, D, and R. Proportion of residue hydrophobic characteristics are presented as decreasing hydrophobicity from 0% up: hydrophobic residue proportions are represented in black, neutral in dark gray, and hydrophilic in light gray.

Figure 6. Mean genetic distances within the core coding region between early and late clonal sets for progressors and nonprogressors. Distance was calculated via the mean Kimura 2-parameter pairwise distances for each patient. At baseline, there were no statistically significant differences in diversity between progressors and nonprogressors. Over time, 5 progressors and 1 nonprogressor showed a decrease in genetic distance within early and late sets of clones. However, mean between-group genetic distances for early and late clonal sets were 0.04 (SE 0.01) and 0.02 (SE 0.003) for progressors vs. nonprogressors, respectively ($p = 0.02$, data not shown). 80% of the progressors had greater overall genetic distance than their matched controls.